CONFIDENCE INTERVAL

(Part 1)

Known standard deviation of population

RANDOM SAMPLING

Conditions to be met:

- a. Every case in the population must have an <u>equal chance</u> of being selected
- b. The selection of a case can in no way <u>affect</u> the selection of any other case

Ex: Alternately selecting a Republican and Democrat violates this condition

c. Cases must be selected in such a way that <u>all</u> <u>combinations are possible</u>

Example: <u>100 marbles in a box</u>; **50 red**, **50 blue**

Blindfolded, you draw 10 marbles randomly

- 1. Each has equal chance
- 2. Selecting red does not affect next marble
- 3. Any combination is possible

SAMPLING ERROR

Population may be large

Sample is representative of population

What is the average age of university students taking six hours or more of coursework?

Population = 10,000 students

Sample = 200 students

Ages range from very young to very old

(Infinite number of sample outcomes)

Mean of sample may not equal mean of population

SAMPLING DISTRIBUTION OF SAMPLE MEANS

Run sample 1000x and calculate/plot means



calculate mean of sample means

standard deviation of sample means

= standard error of the mean

CENTRAL LIMIT THEOREM

"If repeated random samples of size *n* are taken from a population with a mean (μ) and a standard deviation (σ), the sampling distribution of sample means will have a mean equal to μ and a standard error equal to $\sigma/\text{sqrt}(n)$. Moreover, as *n* increases, the sampling distribution will approach a normal distribution."



What is the average age of university students taking six hours or more of coursework?

Cannot assume the sample mean = population mean

however

Might say, "I believe the mean age (μ) of the university is between 23.4 and 26.1 years." based on our sample.

Confidence Interval for the Mean

Confidence Interval for the Mean of population with known σ SAT Scores --- μ = 500, σ = 100

Assume a sample:

 $n = 225 \quad x = 606 \quad (N = 10,000 \quad \mu = ?)$

C.I. = Sample Mean $\pm Z \times (?)$

where Z = 1.96 for 95% confidence interval, or 2.58 for 99% confidence interval and (?) is the Standard Error of the Mean

Review of Z score







C.I. = Sample Mean ± Z ×(?) (?) = σ/\sqrt{n} σ = 100 n = 225 x = 606 Z = 2.58 (for 99%)

 $(?) = 100/\sqrt{225} = 100/15 = 6.67$

C.I. = 606 ± (2.58 x 6.67) C.I. = 606 ± 17.21 C.I. = 588.79 to 623.21 Translation: <u>99</u> times out of <u>100</u> our results would contain the mean of the population

(588.79 - 623.21)

What if we set our confidence level at 95%?

- $CI = 606 \pm 1.96 \times (100/\sqrt{225})$
 - $= 606 \pm 1.96 \times 6.67$
 - = (592.93 to 619.07)

hence, <u>95</u> times out of <u>100</u> our results would contain the mean of the population

CONFIDENCE INTERVAL

(Part 2 of 2)

Unknown standard deviation of population

Confidence Interval for the Mean of population with known σ SAT Scores --- σ = 100

Assume a sample:

 $n = 225 \quad x = 606 \quad (N = 10,000 \quad \mu = ?)$

C.I. = Sample Mean ± Z ×(?)

where Z = 1.96 for 95% confidence interval, or 2.58 for 99% confidence interval and (?) is the Standard Error of the Mean **C.I.** = Sample Mean $\pm Z \times (\sigma/\sqrt{n})$

$\sigma = 100$ n = 225 x = 606 Z = 2.58 (for 99%)

C.I. = $606 \pm (2.58 \times 100/\sqrt{225})$ C.I. = 606 ± 17.21 C.I. = 588.79 to 623.21

Translation: <u>99</u> times out of <u>100</u> our results would contain the mean of the population somewhere between 588.79 – 623.21

Confidence Interval with σ Unknown

Two problems:

- 1. Unknown σ
- 2. Cannot rely on Normal Curve

Use estimate of the standard error of the mean $s_x = sample std dev / square root of n$

$$s_x = s / \sqrt{n}$$

Example: Want to know the average expenditure per customer in the bookstore

Sample size = 100

Mean = \$31.50

Std Dev = \$4.75

$$s_x = s / \sqrt{n} = 4.75 / \sqrt{100} = 4.75 / 10 = .475$$

 $s_x = 0.48$ (Est of std error of mean)

2nd problem: Don't know population mean or std dev, thus cannot use normal distribution or Z value

Not to fear; statisticians to the rescue:

family of *t* distributions

Gossett who worked for Guiness Brewery developed the concept

t distribution :"the shape of a sampling distribution depends on the number of cases in each of the cases"



Small sample size

Larger sample size



Still larger sample size

Similar to Z values

Another concept: **Degrees of Freedom**

In a distribution of *n* cases, *n*-1 cases are free to vary

Example: Quiz worth 10 points

Five scores (n=5), Mean = 8,

Four (*n*-1) of the five are free to vary

Assume four scores are 8, 8, 10, 10

(They could have been anything from 0-10)

Once the four are known, the fifth one cannot vary

(It must be 4)

(5x8=40) & (8+8+10+10=36) therefore 40-36 = 4

Freedom .20 .10 .05 .02 5 1.476 2.015 2.571 3.365	.01 4.032 3.707 3.499	.001 6.869 5.959
5 1.476 2.015 2.571 3.365 1 1 0.042 0.0447 0.142	4.032 3.707 3.499	6.869 5.959
6 1.440 1.042 2.447 2.142	3.707 3.499	5.959
6 1.440 1.943 2.447 3.143	3.499	
7 1.415 1.895 2.365 2.998		5.408
8 1.397 1.860 2.306 2.896	3.355	5.041
9 1.383 1.833 2.262 2.821	3.250	4.781
10 1.372 1.812 2.228 2.764	3.169	4.587
11 1.363 1.796 2.201 2.718	3.106	4.437
12 1.356 1.782 2.179 2.681	3.055	4.318
13 1.350 1.771 2.160 2.650	3.012	4.221
14 1.345 1.761 2.145 2.624	2.977	4.140
15 1.341 1.753 2.131 2.602	2.947	4.073
16 1.337 1.746 2.120 2.583	2.921	4.015
17 1.333 1.740 2.110 2.567	2.898	3.965
18 1.330 1.734 2.101 2.552	2.878	3.922
19 1.328 1.729 2.093 2.539	2.861	10 3.883
20 1.325 1.725 2.086 2.528	2.845	3.850
60 1.296 1.671 2.000 2.390	2.660	3.460
80 1.292 1.664 1.990 2.374	2.639	3.416
100 1.290 1.660 1.984 2.364	2.626	3.390
120 1.289 1.658 1.980 2.358	2.617	3.373
Infinity 1.282 1.645 _ 1.960 2.327	2.576	3.291

Degrees of Freedom

(*n*-1)

Degrees of	LEVEL OF SIGNIFICANCE								
Freedom	.20	.10	.05	.02	.01	.001			
5	1.476	2.015	2.571	3.365	4.032	6.869			
6	1.440	1.943	2.447	3.143	3.707	5.959			
7	1.415	1.895	2.365	2.998	3.499	5.408			
8	1.397	1.860	2.306	2.896	3.355	5.041			
9 1.383 10 1.372	1.383	1.833	2.262	2.821	3.250	4.781			
	1.372	1.812	2.228	2.764	3.169	4.587			
11	1.363	1.796	2.201	2.718	3.106	4.437			
12	1.356	1.782	2.179	2.681	3.055	4.318			
ance 13	1.350	1.771	2.160	2.650	3.012	4.221			
14	1.345	1.761	2.145	2.624	2.977	4.140			
15	1.341	1.753	2.131	2.602	2.947	4.073			
16 1.337	1.746	2.120	2.583	2.921	4.015				
17	1.333	1.740	2.110	2.567	2.898	3.965			
18	1.330	1.734	2.101	2.552	2.878	3.922			
19	1.328	1.729	2.093	2.539	2.861	10 3.883			
20	1.325	1.725	2.086	2.528	2.845	3.850			
60	1.296	1.671	2.000	2.390	2.660	3.460			
80	1.292	1.664	1.990	2.374	2.639	3.416			
100	1.290	1.660	1.984	2.364	2.626	3.390			
120	1.289	1.658	1.980	2.358	2.617	3.373			
Infinity	1.282	1.645	1.960	2.327	2.576	3.291			
	Degrees of Freedom 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 60 80 100 120 Infinity	Degrees of Freedom.2051.47661.44071.41581.39791.383101.372111.363121.356131.350141.345151.341161.337171.333181.330191.328201.296801.2921001.2901201.289Infinity1.282	Degrees of Freedom.20.1051.4762.01561.4401.94371.4151.89581.3971.86091.3831.833101.3721.812111.3631.796121.3561.782131.3501.771141.3451.761151.3411.753161.3371.746171.3331.740181.3301.734191.3281.729201.3251.725601.2961.671801.2921.6641001.2901.6601201.2891.658Infinity1.2821.645	Degrees of Freedom.20.10.0551.4762.0152.57161.4401.9432.44771.4151.8952.36581.3971.8602.30691.3831.8332.262101.5721.8122.228111.3631.7962.201121.3561.7822.179131.3501.7712.160141.3451.7612.145151.3411.7532.131161.3371.7462.120171.3331.7402.110181.3301.7342.101191.3281.7292.093201.2961.6712.000801.2921.6641.9901001.2901.6601.9841201.2891.6581.980Infinity1.2821.6451.960	Degrees of Freedom.20.10.05.0251.4762.0152.5713.36561.4401.9422.4473.14371.4151.8952.3652.99881.3971.8602.3062.89691.3831.8332.2622.821101.3721.8122.2282.764111.3631.7962.2012.718121.3561.7822.1792.681131.3501.7712.1602.650141.3451.7612.1452.624151.3411.7532.1312.602161.3371.7462.1202.583171.3331.7402.1102.567181.3301.7342.1012.552191.3281.7292.0932.539201.3251.7252.0862.528601.2961.6712.0002.390801.2921.6641.9902.3741001.2901.6601.9842.3641201.2891.6581.9802.358Infinity1.2821.6451.9602.327	LEVEL OF SIGNIFICANCEFreedom.20.10.05.02.0151.4762.0152.5713.3654.03261.4401.9422.4473.1433.70771.4151.8952.3652.9983.49981.3971.8602.3062.8963.35591.3831.8332.2622.8213.250101.3721.8122.2282.7643.169111.3631.7962.2012.7183.106121.3561.7822.1792.6813.055131.3501.7712.1602.6503.012141.3451.7612.1452.6242.977151.3411.7532.1312.6022.947161.3371.7462.1202.5832.921171.3331.7402.1102.5522.878191.3281.7292.0932.5392.861201.3251.7252.0862.5282.845601.2961.6712.0002.3902.660801.2921.6641.9902.3742.6391001.2901.6601.9842.3642.6261201.2891.6581.9802.3582.6171nfinity1.2821.6451.9602.3272.576			

Level of Significance

95% Confidence ⁷ Interval (1-.05)

Degrees of	LEVEL OF SIGNIFICANCE							
Freedom	.20	.10	.05	.02	.01	.001		
5	1.476	2.015	2.571	3.365	4.032	6.869		
6	1.440	1.943	2.447	3.143	3.707	5.959		
7	1.415	1.895	2.365	2.998	3.499	5.408		
8	1.397	1.860	2.306	2.896	3.355	5.041		
9	1.383	1.833	2.262	2.821	3.250	4.781		
10	1.372	1.812	2.228	2.764	3.169	4.587		
11	1.363	1.796	2.201	2.718	3.106	4.437		
12	1.356	1.782	2.179	2.681	3.055	4.318		
13	1.350	1.771	2.160	2.650	3.012	4.221		
14	1.345	1.761	2.145	2.624	2.977	4.140		
15	1.341	1.753	2.131	2.602	2.947	4.073		
16	1.337	1.746	2.120	2.583	2.921	4.015		
17	1.333	1.740	2.110	2.567	2.898	3.965		
18	1.330	1.734	2.101	2.552	2.878	3.922		
19	1.328	1.729	2.093	2.539	2.861	3.883		
20	1.325	1.725	2.086	2.528	2.845	3.850		
60	1.296	1.671	2.000	2.390	2.660	3.460		
80	1.292	1.664	1.990	2.374	2.639	3.416		
100	1.290	1.660	1.984	2.364	2.626	3.390		
120	1.289	1.658	1.980	2.358	2.617	3.373		
Infinity	1.282	1.645	1.960	2.327	2.576	3.291		

99% Confidence Interval

(1 - .01)

EXERCISE:

Random sample of 25 retirees. Want to estimate the average number of emails sent out each week. Our sample provides a mean of 12 (emails per week) with a standard deviation of 3. We decide to construct a 95% confidence interval.

CI = Sample Mean ± *t*(s_x)

(Remember: $s_x = s / \sqrt{n}$)

$$CI = 12 \pm t(3/\sqrt{25})$$

	Degrees of	LEVEL OF SIGNIFICANCE							
	(df)	.20	.10	.05	.02	.01	.001		
n = 25	5	1.476	2.015	2.571	3.365	4.032	6.869		
df = 24 (<i>n</i> -1)	7	1.440	1.943	2.447	3.143	3.707	5.959		
$\mathbf{u} = \mathbf{z} + (\mathbf{u} + \mathbf{r})$	8	1.397	1.860	2.306	2.896	3 355	5.408		
	9	1.383	1.833	2.262	2.821	3.250	4.781		
	10	1.372	1.812	2.228	2.764	3.169	4.587		
05% Confidence	11	1.363	1.796	2.201	2.718	3.106	4.437		
	12	1.356	1.782	2.179	2.681	3.055	4.318		
Interval = .05 Level of	13	1.350	1.771	2.160	2.650	3.012	4.221		
Significance	15	1.040	1.761	2.145	2.624	2.977	4.140		
eignneanee	16	1.341	1.753	2.131	2.602	2.947	4.073		
	17	1.333	1.740	2.120	2.583	2.921	4.015		
	18	1.330	1.734	2.101	2.507	2.090	3.965		
t - 2.064	19	1.328	1.729	2.093	2.539	2.861	3.883		
1 - 2.004	20	1.325	1.725	2.086	2.528	2.845	3,850		
	21	1.323	1.721	2.080	2.518	2.831	3.819		
	22	1.321	1.717	2.074	2.508	2.819	3.792		
	23	1.319	1.714	2.069	2.500	2.807	3.768		
	24	1.318	1.711	2.064	2.492	2.797	3.745		
	25	1.316	1.708	2.060	2.485	2.787	3.725		
	20	1.315	1.706	2.056	2.479	2.779	3.707		
	28	1.313	1 701	2.052	2.473	2.771	3.690		
	29	1.311	1.699	2.045	2.462	2.705	3 6 5 9		
	30	1.310	1.697	2.042	2 457	2 750	3 646		
	40	1.303	1.684	2.021	2.423	2.704	3 551		
	50	1.299	1.676	2.009	2.403	2.678	3.496		
	60	1.296	1.671	2.000	2.390	2.660	3.460		
	80	1.292	1.664	1.990	2.374	2.639	3.416		
	100	1.290	1.660	1.984	2.364	2.626	3.390		
	120	1.289	1.658	1.980	2.358	2.617	3.373		
	~	1.282	1.645	1.960	2.327	2.576	3.291		

CI = Sample Mean $\pm t(s / \sqrt{n})$

CI = $12 \pm 2.06(3/\sqrt{25})$ CI = $12 \pm 2.06(3/5)$ CI = $12 \pm 2.06(.6)$ CI = 12 ± 1.24 CI = 10.76 to 13.24

Our population mean is between **10.76** & **13.24**. And, our method will produce a <u>correct estimate</u> <u>95 out of 100 times</u>.

	Degrees of	LEVEL OF SIGNIFICANCE							
	(df)	.20	.10	.05	.02	.01	.(
<i>n</i> = 25	5	1.476	2.015	2.571	3.365	4.032	6.		
<i>df</i> = 24 (<i>n</i> -1)	7	1.415	1.943	2.447 2.365 2.306	3.143 2.998	3.707 3.499	5. 5.		
	9 10	1.383	1.833	2.262	2.890	3.355	5. 4.		
99% Confidence	10 11 12	1.363 1.356	1.796 1.782	2.228 2.201 2.179	2.764 2.718 2.681	3.169 3.106 3.055	4. 4.		
Interval = .01 Level of	13 14	1.350 1.345	1.771 1.761	2.175 2.160 2.145	2.650 2.624	3.033 3.012 2.977	4. 4. 4		
Significance	15 16	1.341 1.337	1.753 1.746	2.131 2.120	2.602 2.583	2.947 2.921	4.0		
	17 18	1.333 1.330	1.740 1.734	2.110 2.101	2.567 2.552	2.898 2.878	3.9 3.9		
<i>t</i> = 2.797 (or 2.80)	19 20	1.328	1.729 1.725	2.093 2.086	2.539 2.528	2.861 2.845	3.8 3.8		
	21 22 23	1.323 1.321 1.210	1.721 1.717	2.080 2.074	2.518 2.508	2.831 2.819	3.8 3.7		
	23	1.319	1.714 1.711	2.069	2.500 2.492	2.807 2.797	3.7 3.7		
	25 26 27	1.316 1.315 1.314	1.708 1.706 1.703	2.060 2.056 2.052	2.485 2.479 2.473	2.787 2.779 2.771	3.7 3.7 3.6		
	28	1.313	1.701 1.699	2.048 2.045	2.467 2.462	2.763 2.756	3.6 3.6		
	30 40 50	1.310 1.303 1.299	1.697 1.684 1.676	2.042 2.021 2.009	2.457 2.423 2.403	2.750 2.704 2.678	3.6 3.5 3.4		
	80	1.296	1.671 1.664	2.000 1.990	2.390 2.374	2.660 2.639	3.4 3.4		
	120 ∞	1.290 1.289 1.282	1.658 1.645	1.984 1.980 1.960	2.364 2.358 2.327	2.626 2.617 2.576	3.3 3.3 3.2		
	MATCRE PROFILE					7 H.			

CI = Sample Mean $\pm t(s / \sqrt{n})$

CI = $12 \pm 2.80(3/\sqrt{25})$ CI = $12 \pm 2.80(3/5)$ CI = $12 \pm 2.80(.6)$ CI = 12 ± 1.68 CI = 10.32 to 13.68

Our population mean is between **10.32 & 13.68**. And, our method will produce a <u>correct estimate</u> <u>99 out of 100 times</u>. Final note:

A confidence interval for the mean <u>does not</u> provide you with an exact estimate of the population mean.

Rather, it provides you with an interval that you believe contains the true mean,

and you are confident that 95% (*or 99%, etc*) of your confidence intervals would contain the true mean.